May 2024

Reweighting Method for Lund String Breaks in PYTHIA

Hugo Engman

Theoretical Particle Physics
Division of Particle and Nuclear Physics
Department of Physics
Lund University

Bachelor thesis (15 hp) supervised by Christian Bierlich

LUND
UNIVERSITY

Abstract

Event generators are useful for simulating collision experiments in high-energy particle physics. In the event generator PYTHIA 8, parameters may be varied to compare competing models against experimental data. It is then beneficial to employ reweighting techniques to explore the results for multiple parameter values with only one simulation.

This study aims to develop and implement a reweighting method for meson production in electron-positron collisions, within the Lund string model for hadronization. We expect the number of $s\bar{s}$, $u\bar{u}$ and $d\bar{d}$ string breaks to follow a multinomial distribution and the produced mesons to distribute correspondingly. Thus, a ratio between two multinomial mass functions using the string break probabilities for two different sets of parameters is developed and used as statistical weight for each event. By varying the $s\bar{s}$ suppression parameter as well as η and η' rejection parameters the weight for each event is calculated and applied to the distribution of final $s\bar{s}$ breaks and mesons from a set of test simulations. The reweighted test distributions are compared to target distributions for comparison.

The results show a high accuracy of the method when applied to the number of string breaks but much lower for the number of final mesons, pointing towards a discrepancy between the predicted correlation of string breaks and mesons and the true correlation.

The reweighting technique introduced can be naturally extended to reweighting around the mixing angles of the pseudoscalar and vector mesons as well as the vector-to-pseudoscalar suppression factor also present. It can also be further generalised to include baryons and account for hadron decays. In the end, one hopes to use this method for comparison with experimental data.

Popular Science Summary

The **event generator** is a powerful tool in particle physics which simulates particle collisions. The event generators generate much simulated data. To reduce the generated data, techniques can be implemented that enable us to take shortcuts in the computing process. One of these shortcuts is **reweighting**, which makes a theoretical prediction of how varying input parameters would change the results. If this is possible, one only has to run the simulation once and then simply reweight the results to explore different sets of outcomes.

Say you want to simulate a collision between two high-energy particles just as it is done at CERN's large accelerators. You then need to know how the particles behave before they reach the detectors. This is the job of event generators, which use physics models implemented as computer algorithms to predict exactly how a particle collision would behave in real life. Then, you want to run more simulations to obtain more data. You may want to do one hundred, or why not a million, or even a hundred million simulations. This is such that the amount of data from the simulations is comparable to the huge amount from the real experiments. Now consider, you have two different values for an input parameter you would like to compare. You would then have to produce two hundred million simulations. What if you want to vary fifty parameters, each over a thousand values? It is easy to see how this will become incredibly time, energy and money-consuming. By narrowing down on one part of the simulation known as **hadronization**, we can simplify the problem.

Hadronization is the process of constituent particles known as quarks turning into composite particles called hadrons. Picture two quarks connected by a string. If the string has enough energy, it can break in two, producing a quark in each new string end. The breaking can continue until the energy of the string is depleted. The quarks in each string piece then combine to form particles known as mesons. This is the **Lund String model** of hadronization.

The quarks produced in the string breaks are of three types: up, down and strange, each with an associated probability of being produced. This can be directly linked to the probability of producing a certain number of hadrons in a simulation. Modifying the individual quark probabilities by altering the input parameters, we can predict how the number of hadrons will change and adjust the result accordingly. Instead of running a new simulation for each parameter value, the original simulation can instead be used to obtain all the results. This is the basis of the reweighting method used in the thesis.

Contents

1 Introduction

Particle accelerators such as CERN's Large Hadron Collider (LHC) are used to accelerate particles and then collide them. In collision experiments, the characteristics of the fundamental particles are probed. To construct such an experiment, a good starting point is a prediction of the outcome. This can be made by combining theoretical descriptions of the individual particles with phenomenological models that describe the collective and dynamical behaviour of many-particle systems. A set of numerical algorithms that aim to simulate the process is known as an event generator. One can then use the simulation to e.g. gain insight into potential experiments. Due to the broad scope of event generation, many parameters and models are needed to describe a full event accurately, making them inherently computationally demanding.

Reweighting is a tool that aims to reduce the number of simulations by adding statistical weights to the results. These weights correspond to the predicted change in the final distribution obtained by varying input parameters. Reweighting allows the exploration of results from multiple different parameter values with only one simulation. In this study, a reweighting technique is developed for the event generator PYTHIA 8 [1], and its associated hadronization model: the Lund string.

The Lund string model [2] describes the outgoing quarks from a collision as attached via strings that fragment iteratively, creating quark-antiquark pairs which combine into hadrons. This is a low-energy phenomenon and thus only allows for light quark production. Disregarding baryons we can limit ourselves to three types of string breaks: $u\bar{u}, d\bar{d}$ and $s\bar{s}$.

The goal is to first introduce a method for reweighting the number of final state mesons resulting from modifying the flavour selection of the string breaks. Specifically, the reweighting pertains to varying the $s\bar{s}$ break suppression parameter ρ arising from the larger relative mass of the s quark. Another reweighting method is then developed to account for acceptance factors $p_{\eta,A}$ and $p_{\eta',A}$, influencing the production of the final state η and η' mesons respectively, as they are further suppressed due to their masses.

Implementing a reweighting method considering these three parameters is a building block for further modification to increase its scope. This project aims to provide a simple but powerful technique for reweighting, to be used and further developed within PYTHIA. As the methods used in this project are not restricted to particle physics, they could also be abstracted and applied to statistical simulations in other fields.

The thesis will give a brief introduction of event generation and the Lund String model. Then it will describe the physical concepts related to string breaks and meson selection, and state the resulting parameters used for hadronization in PYTHIA. In the method, an initial binomial distribution is assumed for reweighting string breaks which is then

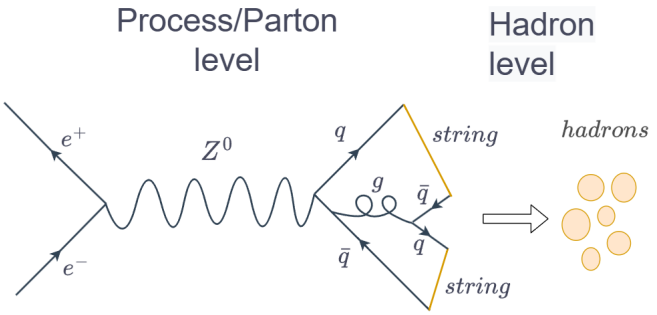

Figure 1: Feynman diagram of e^+e^- collision resulting in string formation ending with final hadrons with the highlighted Hadron level processes being the focus of the thesis.

expanded to a multinomial distribution by including the rejection of η and η'. The accuracy of the reweighting method is measured in the results and a conclusion of the project is given.

2 Theory

To begin reweighting at hadronization level, it is beneficial to look at a collision where string fragmentation is the main process and a low number of strings is produced. Therefore, the following interaction is examined:

$$e^+e^- \to Z^0 \to q\bar{q},$$

shown in figure 1. This section will start by describing the fundamentals of event generation before narrowing down on the physics behind hadronization and the Lund string. This outlines the reason behind the input parameters used in PYTHIA for flavour selection and meson production. Lastly, a table of the used parameters and an example of the steps to produce a meson is given.

2.1 Event generators

Due to the quantum nature of a collision event, randomness is an inherent characteristic of the results. Thus, event generation needs to have the same approach through random number generation and integration techniques which give the programs the name Monte Carlo event generators [1]. The output of the event generators represents an ideal detec-

tor that captures key particle properties where only the number of produced particles is considered here.

The events are structured to evolve in a transverse momentum ordered or 'hardness' scale. This means that the hardest processes with the largest transverse momentum are calculated first, and the system evolves from there. A simplified picture of the process can then be ordered accordingly[1]:

$$\text{Process level} \rightarrow \text{Parton level} \rightarrow \text{Hadron level},$$

where the process level handles the hardest processes such as the initial parton collision and the resulting short-lived resonances. At parton level, the parton showers are generated and at hadron level, hadronization and hadron decay are handled. These parts are considered in the context of the electron-positron collision in figure 1. This study considers the formation of strings to the right in the figure and how they evolve to form hadrons i.e. hadronization.

2.2 Hadronization and the Lund String

In high-energy interactions, the characteristics of QCD are of high interest. In the initial hard processes, perturbative QCD can be used. As the event evolves and the energy scale decreases, this is no longer possible, and new models are introduced. One of the models for hadronization is the Lund String model which captures the underlying structure and dynamics of QCD while also describing the more diffuse, collective behaviour of many particles [3].

The Lund model describes a massless, relativistic string between quarks acting as the QCD force field with string tension $\kappa = 1$ GeV/fm determining the equations of motion [2]. The semi-classical model describes hadronization through stochastic string breaks with quarks and gluons forming the excitations at the ends and interior of the strings respectively, as well as tunnelling processes resulting in new quark-antiquark pairs.

2.3 String decay

The main property of the Lund string considered for this thesis is that it can decay through fragmentation. For meson production, this is done by producing a new $q\bar{q}$ pair at a single vertex in the interior of the string, splitting it into two colour singlet string pieces. This process repeats until the energy necessary for the production process is depleted. What remains are the final state mesons. This is shown in figure 2.

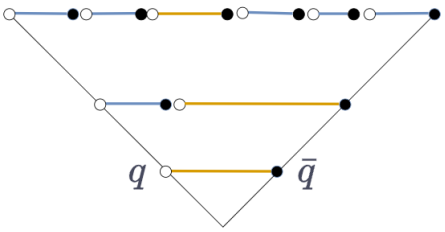

Figure 2: Evolution of a fragmenting string

2.3.1 String breaks

The fragmentation function [3], describes an iterative process of the string breaks as a function of the momentum fraction z of the remaining string momentum. Requiring the string to fragment on either side of the centre with equal probability, one gets a symmetric probability distribution of a quark-antiquark pair of flavours α and β combining as

$$f_{\alpha\beta}(z) \propto N_{ab}\frac{1}{z}z^{a_\alpha}\left(\frac{1-z}{z}\right)^{a_\beta}\exp\left(\frac{-bm^2}{z}\right), \tag{1}$$

where m is the mass of the produced meson. a and b are parameters, the first is flavour-dependent while the second is not. The function can be simplified by allowing a to be the same for all flavours, yielding a flavour-independent fragmentation

$$f(z) \propto \frac{(1-z)^a}{z}\exp\left(\frac{-bm^2}{z}\right), \tag{2}$$

which is the one implemented in PYTHIA for the production of mesons [1].

2.3.2 Flavor Selection

The property of interest in this thesis is the flavour selection process in the string breaks which must be determined quantum mechanically. Classically, the linearity of the string requires a massless quark-antiquark pair to be produced in a single vertex with no transverse momentum to preserve the energy and momentum of the system. If the quarks were to have mass or transverse momentum, they would have to be produced at a distance l from the break point such that the energy stored in the field κl is large enough to account for the newly produced transverse mass $m_\perp$:

$$m_\perp^2 \equiv m^2 + p_\perp^2. \tag{3}$$

To achieve this, the quarks must tunnel through the classically forbidden region $|x| < l$. To describe the wave function in this region, one may use the WKB approximation to obtain the assumed solution for a slowly varying potential with a transmission coefficient T given by

$$T \propto e^{-2\int |p(x)|dx} = \exp\left(-\frac{\pi m^2}{\kappa}\right), \tag{4}$$

as derived in Appendix A.

Allowing $m \to m_\perp$, this results in a probability distribution for producing a pair of quarks with equal but opposite transverse momentum, this is then factorised as [4]

$$P(m^2, p_\perp^2) \propto \exp\left(-\frac{\pi m_q^2}{\kappa}\right) \exp\left(-\frac{\pi p_\perp^2}{\kappa}\right). \tag{5}$$

The first term introduces a suppression of the production of heavier quarks, estimated using constituent quark masses to be [3]

$$u : d : s : c \approx 1 : 1 : 1/3 : 10^{-11}. \tag{6}$$

Hence, we can restrict the quarks produced in string breaks to up, down and strange. It is not a given what values to assign to the s-quark mass m_s, and m_{ud}: the u,d-quark masses. Thus the strange suppression parameter $\rho = -\exp\left(\frac{\pi(m_s^2 - m_{ud}^2)}{\kappa}\right)$ is left as a free parameter, and in PYTHIA it has the name `StringFlav:probStoUD` and is by default set to 0.217 [1].

2.4 From Quarks to Hadrons

After selecting the quark flavours, two other properties are considered when choosing meson type: spin and mixing.

2.4.1 Vector Meson suppression

Restricting the mesons to ground state $L = 0$ leaves two spin-states: $S = 1, 0$ mesons. The vector mesons have spin $S = 1$. Due to this, they form three states with projections

$m_S = 1, 0, -1$. Comparing this to the pseudoscalar mesons with $S = 0$, one could expect the ratio between them to be 3:1. This is however not the case and vector mesons are in actuality produced in smaller quantities than the pseudoscalars. This is due to the vector mesons' higher mass, caused by the internal spin-spin interactions between the quarks. The mass of a meson can be modelled as [2]

$$M = m_1 + m_2 + A\frac{\mathbf{S}_1 \cdot \mathbf{S}_2}{m_1 m_2}, \tag{7}$$

where $\mathbf{S}_{1,2}$ is the spin of the quark and antiquark, $m_{1,2}$ are the masses and A is a proportionality factor.

The spin-spin interaction is negative for pseudoscalars and positive for vectors [2]

$$\mathbf{S}_1 \cdot \mathbf{S}_2 = \begin{cases} \frac{1}{4}\hbar^2 \text{ for vectors.} \\ \frac{-3}{4}\hbar^2 \text{ for pseudoscalars.} \end{cases} \tag{8}$$

The pseudoscalars thus have lower mass and are enhanced to minimise the energy of the state. PYTHIA handles this by introducing suppression parameters `StringFlav:mesonUDvector` = 0.50 for up/down mesons and `StringFlav:mesonSvector` = 0.55 for strange mesons, yielding vector meson probabilities 0.5/(1+0.5) and 0.5/(1+0.55) respectively.

2.4.2 Mixing

By combining all quark-antiquark pairs of u, d and s, two types of configurations arise, the pseudoscalar nonet and the vector nonet of mesons [5]. Two of the pseudoscalar states have identical quantum numbers and similar quark composition, allowing them to mix: the η and η' mesons. The two η flavour states are

$$\begin{aligned} \eta_8 &= (u\bar{u} + d\bar{d} - 2s\bar{s})/\sqrt{6} \\ \eta_1 &= (u\bar{u} + d\bar{d} + s\bar{s})/\sqrt{3} \end{aligned} \tag{9}$$

where η_1 relates to the singlet state of the nonet and η_8, the octet state.

The observed η and η' states are formed by mixing these two states by an (orthogonal) linear combination given by [5]

$$\begin{aligned} \eta &= \cos(\theta_\eta)|\eta_0\rangle + \sin(\theta_\eta)|\eta_8\rangle \\ \eta' &= -\sin(\theta_\eta)|\eta_0\rangle + \cos(\theta_\eta)|\eta_8\rangle \end{aligned} \tag{10}$$

Table 1: Table of parameters used for meson production

Name	Value
(ρ) StringFlav:probStoUD	0.217
StringFlav:mesonUDvector	0.50
StringFlav:mesonSvector	0.55
StringFlav:thetaPS	-15 °
($p_{\eta,A}$)StringFlav:etaSup	0.6
($p_{\eta',A}$)StringFlav:etaPrimeSup	0.12

In PYTHIA, the mixing is introduced as StringFlav:thetaPS = $-15°$. It is then converted into angle $\alpha = 90° - (\texttt{StringFlav:thetaPS} + 54.7°)$: the mixing angle between the flavour states of the s and u/d quarks [6]. The mixing probability is given as $\cos^2\alpha$ for η and $\sin^2\alpha$ for η' for an $s\bar{s}$ meson. Given a $u\bar{u}$ or $d\bar{d}$ meson, there is a probability of 0.5 of a π^0, $\frac{\sin^2\alpha}{2}$ of η and a $\frac{\cos^2\alpha}{2}$ of η'.

This choice overestimates the production of η and η'. To account for this, the acceptance parameters StringFlav:etaSup and StringFlav:etaPrimeSup are then introduced with values of 0.6 and 0.12 respectively.

The rest of the lighter ($S = 0, 1$) strange mesons considered consist of the pseudoscalar kaons

$$K^+ = u\bar{s}, \quad K_L^0 = \frac{d\bar{s} - \bar{d}s}{\sqrt{2}}, \quad K_S^0 = \frac{d\bar{s} + \bar{d}s}{\sqrt{2}}, \quad K^- = s\bar{u} \tag{11}$$

and the vectors.

$$\phi = s\bar{s} + \text{ small mixing angle} \tag{12}$$

A table summarising the PYTHIA parameters used and their default value is shown in table 1.

An example of an η meson produced from string break is shown in figure 3 and can be summarised as follows: Start with an s quark. $\rightarrow$ Fragment and pick $s\bar{s}$ string break with probability $p_{ss} = \frac{\rho}{2+\rho}$. $\rightarrow$ Pick it to be spin 0 with probability $p_{s,J=0} = 1/(1+0.55)$. $\rightarrow$ Choose η with probability $\cos^2(\alpha)$. $\rightarrow$ Accept it with probability $p_{\eta,A} = 0.6$. The full meson selection and associated probabilities are considered in the method.

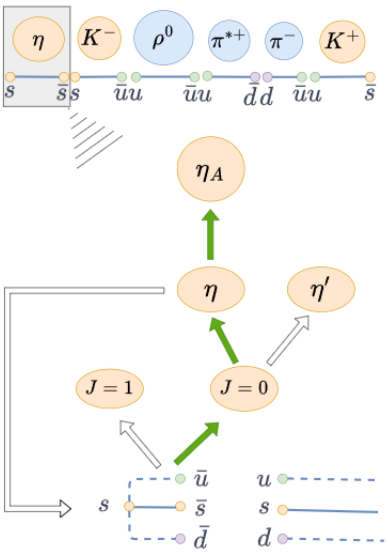

Figure 3: (Top) final string breaks and resulting hadrons: yellow if strange, blue if not. (Bottom) the selection of an η meson from a set of choices based on PYTHIA parameters.

3 Method

The strategy is to first consider the simple case of a binomial reweighting of the ρ parameter from flavour selection. First passage probabilities [7] are introduced to calculate weights when rejection is considered. To incorporate the η and η' rejection, an updated multinomial reweighting method includes terms for meson production. State diagrams, showing transition probabilities are used as visual tools.

3.1 Reweighting a binomial

The choice between a $s\bar{s}$ break and a non $s\bar{s}$ break is a trial with two outcomes: success (s) or fail (ud). It is illustrative to consider a similar case of coin tossing. One can use the binomial distribution [8] to predict the number of heads, r, of N throws with a given probability p to get heads:

$$\binom{N}{r} p^r (1-p)^{N-r}, \tag{13}$$

The law of large numbers tells us that by repeating independent and identical samples

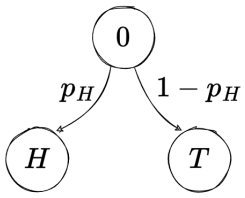

Figure 4: State diagram representing a coin toss. Starting from a state 0, there is a probability of p_H and $1 - p_H$ of reaching state H (head) and T (tail) respectively.

many times, the sample distribution will converge to the parent distribution. In this case, it is then true that for a given N, and two coins with different probabilities of head p_1 and p_2, one can calculate the relative likelihood of an outcome r by taking the ratio of the theoretical values obtained in eq. (13):

$$\left(\frac{p_2}{p_1}\right)^r \cdot \left(\frac{1-p_2}{1-p_1}\right)^{N-r}. \tag{14}$$

Sampling the number of heads using the coin associated with p_1, it is then possible to multiply it by the factor to approximate the coin associated with p_2. Hence, we have our binomial reweighting factor.

It is then possible to use this factor by relating p to the probability of an $s\bar{s}$ break, N to the total number of breaks and r to the number of $s\bar{s}$ breaks.

3.2 Reweighting a biased coin

The first step is to implement reweighting for the simplest case, a binomial distribution from a coin toss. Sampling random numbers from a uniform distribution between 0 and 1, this was implemented as a simple example of the reweighting factor introduced in eq. (14) applied to an initially unbiased coin.

An illustrative tool is a state diagram, typically applied to Markov Chains, defined as processes where probabilities of the next possible outcomes depend only on the current state [7] as shown in figure 4. Some results of the reweighting are displayed in figure 5. The histograms highlight some of the features of reweighting. Firstly, the precision and accuracy decrease in the fringes where there are fewer samples. Secondly, as the distributions converge, the reweighting becomes more exact, and the range of values it captures increases.

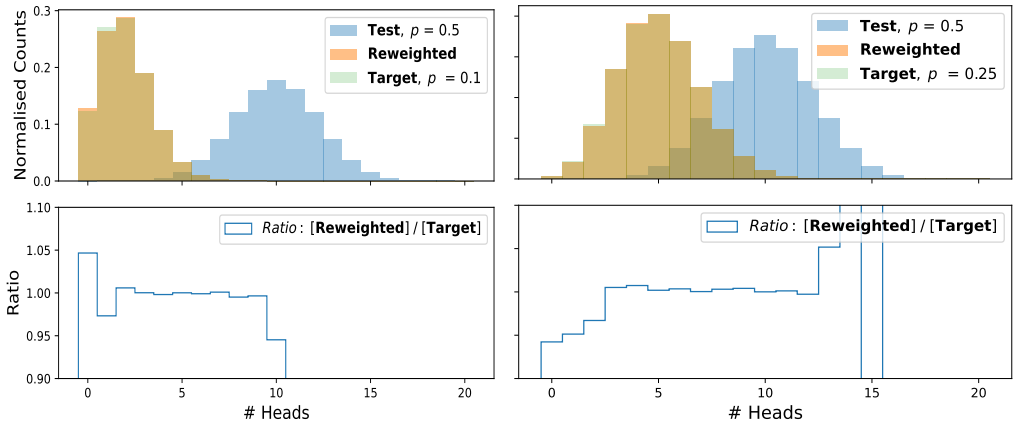

Figure 5: Distribution of heads of twenty coin-tosses reweighted from $p = 0.5$ to different values.

3.3 Reweighting Hadrons

To reweight the hadrons at string break level, the state diagram in figure 6 is first considered. The total probability of an event is broken down into the probability of choosing r strange breaks out of N strings and then picking k desired strange mesons out of them

$$\binom{N}{r} p_S^r (1-p_S)^{N-r} \binom{r}{k} p_m^k (1-p_m)^{r-k}, \tag{15}$$

where p_m is the probability of obtaining the meson. Since p_m is not altered when reweighting ρ, only p_S needs to be considered and we get back eq. (14).

3.4 Reweighting a biased coin with rejection

The next step is to reweight the rejection of the η and η' mesons. Rejection is another form of suppression where instead of lowering the probability, rejecting the outcome and starting over from the initial state is possible. In the case of coin tossing, if one introduces a rejection probability when tossing a head, there is a possibility when flipping a head that we have to flip the coin again. This is represented in the diagram of figure 7. The binomial reweighting formula is no longer valid as the probability of obtaining a head is modified. The new probability can be calculated with the concept of first passage probability.

To describe the probability of reaching one state in a Markov chain before another, one

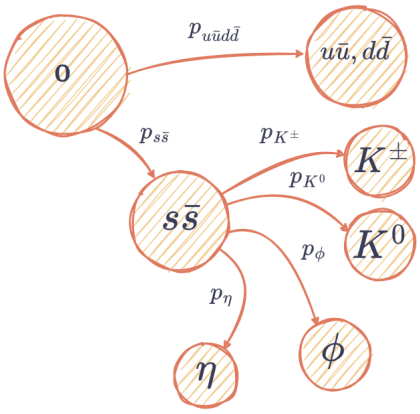

Figure 6: State diagram indicating the transition probabilities from string break (0) to strange mesons.

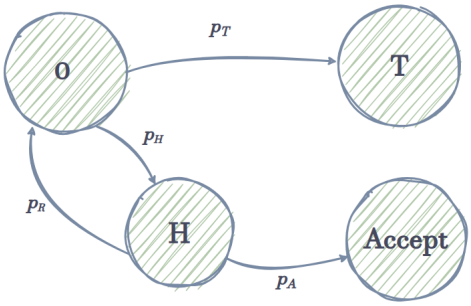

Figure 7: Accept Reject algorithm, coin toss.

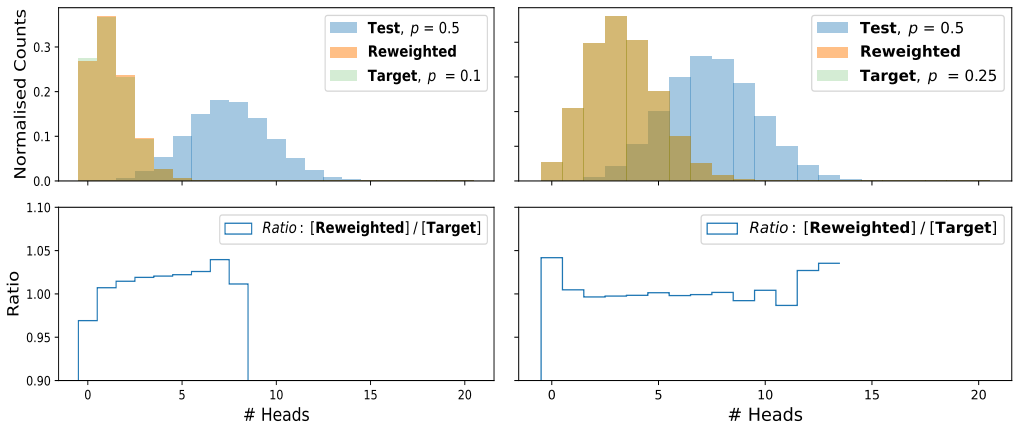

Figure 8: Distribution of heads of twenty coin-tosses with an acceptance rate of 0.6, reweighted from $p = 0.5$ to different values.

can use the concept of first passage time, defined as $T_A = \min\{n \geq 0 : X_n \in A\}$. It can then be proven that $h_i = P(T_A < T_B | X_0 = i)$, that is, the probability of reaching state A before B is [7]

$$h_i = \begin{cases} 0 & \text{if } i \in B \\ 1 & \text{if } i \in A \\ \sum_{j \in S} p_{ij} h_j & \text{if } i \in S - (A \cup B) \end{cases} \tag{16}$$

where p_{ij} is the one-step probability of going from i to j. This 'algorithm' will give us a system of equations to solve for the probability of heads from the initial state. A is the state Accept and B is state T, the probability of throwing a coin and ending with heads is then

$$h_0 = p_H \cdot p_A + p_H \cdot p_R \cdot h_0 \tag{17}$$

$$\iff h_0 = \frac{p_H p_A}{1 - p_H p_R}$$

An example distribution of the reweighting is again shown in figure 8. This plot highlights the statistical uncertainty which arises when the differences between the two distributions increase.

3.5 Reweighting hadrons with rejection

Including the rejection of η and η' introduces three difficulties. First, the rejection probability changes depending on the other quark type in the forming meson. Second, $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ breaks all have to be considered individually as η and η' can form from both $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ mesons. Therefore, a multinomial probability has to be introduced. Third, the rejection occurs in the meson-choosing process which complicates the probability calculations and reduces the simplicity of the method. State diagrams showing a string piece either with an s or u/d in one end and fragmenting into $s\bar{s}$ or $u\bar{u}, d\bar{d}$ in the other are shown in figures 9,10.

The diagrams show the total string-break selection from parameters discussed in the theory section:

1. Start with an initial s or u/d quark
2. Pick the next string break flavour with probabilities $\frac{\rho}{2+\rho}, \frac{1}{2+\rho}, \frac{1}{2+\rho}$
3. If same as initial flavour, probability $p_{s,J=0}$ or $p_{ud,J=0}$ of picking a pseudoscalar meson depending on the flavour.
4. If pseudoscalar, pick light, diagonal meson based on mixing angle α.
5. If η or η' accept with probability $p_{\eta,A}$ or $p_{\eta',A}$ respectively.

The start quark to end quark probabilities are divided into five unique cases depending on the initial and final quark flavour: $p_{s,s}$,$p_{s,ud}$,$p_{ud,s}$,$p_{ud,ud}$,$p_{ud,du}$. The probability of an event is given as a multinomial probability mass function [8]

$$\frac{N!}{N_{s,s}!N_{s,ud}!N_{ud,s}!N_{ud,ud}!N_{ud,du}!} p_{s,s}^{N_{s,s}} p_{s,ud}^{N_{s,ud}} p_{ud,s}^{N_{ud,s}} p_{ud,ud}^{N_{ud,ud}} p_{ud,du}^{N_{ud,du}} \tag{18}$$

altering the probabilities in eq. (18) and dividing, the multinomial reweighting factor is:

$$\left(\frac{p_{2s,s}}{p_{1s,s}}\right)^{N_{s,s}} \left(\frac{p_{2s,ud}}{p_{1s,ud}}\right)^{N_{s,ud}} \left(\frac{p_{2ud,s}}{p_{1ud,s}}\right)^{N_{ud,s}} \left(\frac{p_{2ud,ud}}{p_{1ud,ud}}\right)^{N_{ud,ud}} \left(\frac{p_{2ud,du}}{p_{1ud,du}}\right)^{N_{ud,du}} \tag{19}$$

The probabilities can be determined using the first passage method as in eq. (17).

$$p_{s,s} = 1 - p_{s,ud} \tag{20}$$

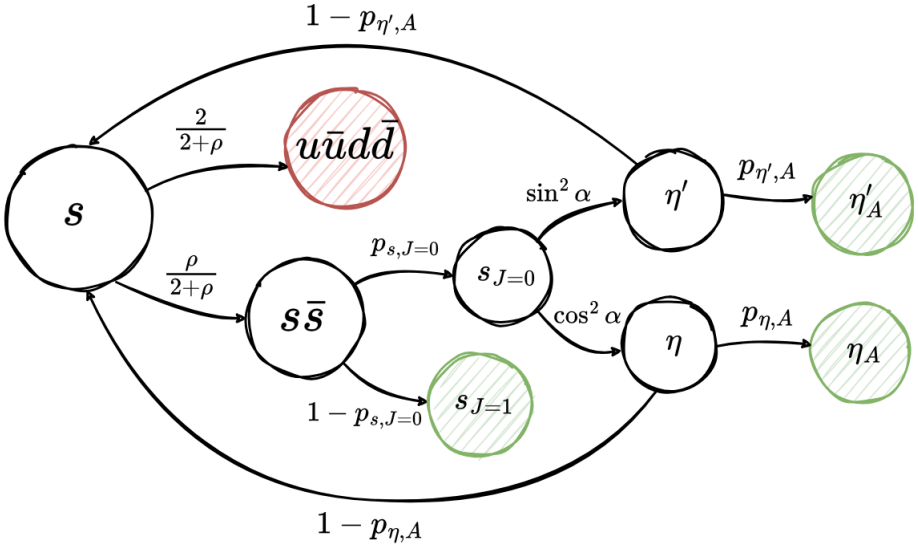

Figure 9: Transition probabilities starting from an initial s quark ending at states either with an $s\bar{s}$ break (green) or $u\bar{u}d\bar{d}$ break (red).

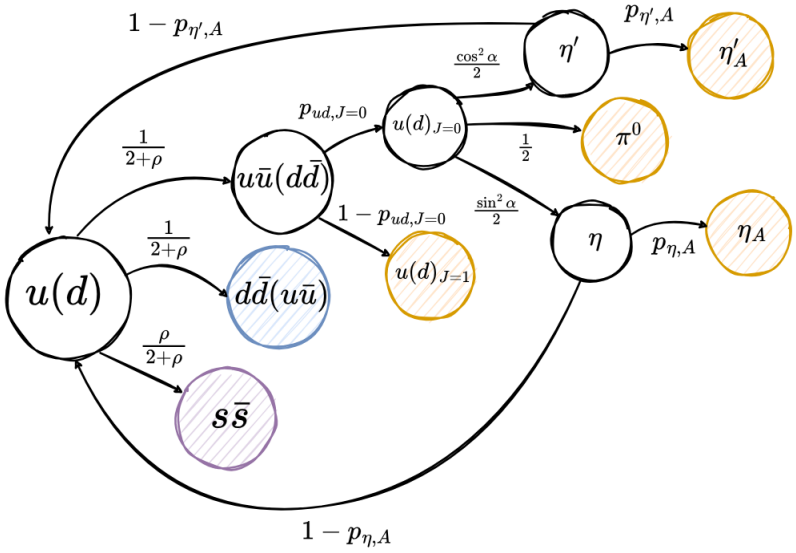

Figure 10: Transition probabilities starting from an initial $u(d)$ quark ending at states either with an $s\bar{s}$ break (purple), $u\bar{u}(d\bar{d})$ break (yellow) or $d\bar{d}(u\bar{u})$ break (blue).

$$p_{s,ud} = \frac{2}{2+\rho(1-p_{s,J=0}((1-p_{\eta,A})\cos^2\alpha+(1-p_{\eta',A})\sin^2\alpha))} \tag{21}$$

$$p_{ud,s} = \frac{\rho}{2+\rho-\frac{p_{ud,J=0}}{2}((1-p_{\eta,A})\sin^2\alpha+(1-p_{\eta',A})\cos^2\alpha)} \tag{22}$$

$$p_{ud,du} = \frac{1}{2+\rho-\frac{p_{ud,J=0}}{2}((1-p_{\eta,A})\sin^2\alpha+(1-p_{\eta',A})\cos^2\alpha)} \tag{23}$$

$$p_{ud,ud} = 1 - p_{ud,s} - p_{ud,du} \tag{24}$$

The full derivation, based on the state diagrams, is shown in Appendix B.

3.6 Implementation

The main part of the thesis consisted of developing the reweighting method and use the formalism described in the method to calculate the analytic expressions for the reweighting factors. This was then implemented in PYTHIA with the following base setup

- Disable baryon production, hadron decays and allow Z^0 to decay only to light quark-antiquark pairs.
- Limit gluon decay in the parton shower to light quarks as well.
- Set e^+e^- beams COM energy to 91.18 GeV$/c^2 = Z^0$ mass and limit hard process to Z^0 production.
- Generate events with standard parameter values, collecting the number of string breaks given their neighbour ($N_{s,s}, N_{s,ud}N_{ud,s}N_{ud,ud}N_{ud,du}$) and storing the number of desired particles in a 'test' histogram.
- Calculate weights using eq.(19) from the number of string breaks and new input values for one of the ρ, p_{η_A} and $p_{\eta'_A}$ parameters.
- Apply weights at the observable level to desired particles, constructing a 'reweighted' histogram.
- Generate events with input parameters, storing the number of desired particles in 'target' histogram
- Show the three histograms in the same figure, comparing the 'reweighted' and 'target' distribution using a ratio plot and goodness-of-fit test.
- Repeat for the other two parameters

4 Results

Plots showing the reweighted distributions are displayed. ρ is initially reweighted using the binomial formula, then the multinomial one for comparison, followed by reweighting η_A and η'_A. The number of $s\bar{s}$ breaks is first considered as the number of string breaks is the property being used for reweighting and provides insight into the overall accuracy of the method.

K^+, K^- are then considered along with η, η' for ρ reweighting. Switching to η_A, η'_A reweighting, the η, η' plots are changed to all diagonal mesons ($u\bar{u}$,$d\bar{d}$,$s\bar{s}$). The meson plots show if the reweighting method is exact along with the assumptions made for the string break to meson transition. These are the most interesting and useful plots as it is the mesons that are measured in real experiments.

The ratio plots highlight the qualitative difference between the reweighted and target distribution compared to the difference between the test and target distribution. A Pearson chi-squared test [8] is performed for each plot and a p-value is calculated which gives the probability that the reweighted distribution is sampled from the target distribution. This is not used as an absolute measure of precision or hypothesis testing. It is instead used as a quantitative tool for relative comparison between plots: a higher p-value signifies a closer relationship between the reweighted distribution and the target one.

4.1 Binomial ρ Reweighting

Plots showing the initial ρ reweighting method considering the binomial selection without rejection from eq. (14) are shown in figure 11. The top plots show the number of $s\bar{s}$ string breaks, the middle ones show charged kaons (K^+, K^-) and the bottom one η, η' mesons. They are each reweighted for a 1% and 10% decrease in ρ. Some key features are:

The η, η' production does not vary much when altering ρ. This can be seen as the [**Test**]/[**Target**] stays close to one. The reduced gap between the two distributions increases the p-value as the reweighted distribution is also inherently close to the target. In the η, η' histograms there are also larger error bars towards larger N meaning low counts and more simulations might be necessary. Comparing the ratio plots [**Reweighted**]/[**Target**] for 1% reweighting, they all seem to be very close to 1. However, comparing the p-values ($s\bar{s} \approx 10^{-1}; K^{\pm} \approx 0; \eta, \eta' \approx 10^{-3}$) highlights the accuracy of reweighting the number of $s\bar{s}$ breaks compared to mesons. Looking at the 10% reweighting, the ratio plots start to diverge and this is reflected in the p-values, all tending to 0.

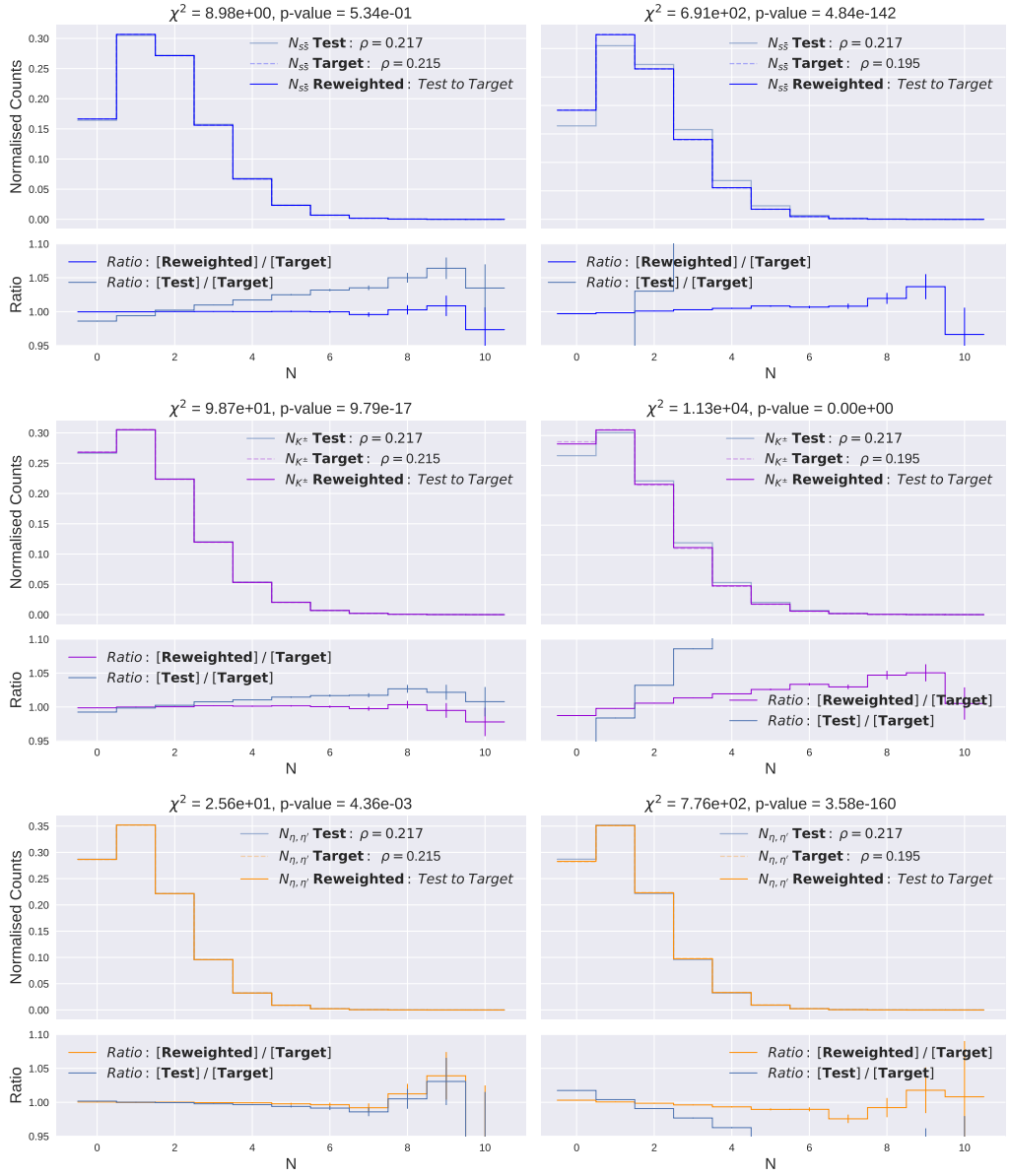

Figure 11: Six normalised histogram plots, showing Test, Target and Reweighted histograms of final particles of 10^8 events, accompanied by plots showing the ratio between them to highlight their relative difference. On the top of each plot is a Pearson's chi-squared total test statistic and its associated p-value used for internal comparison. Error bars show normalised poison error on the mean for each N value. ρ is reweighted from 0.217 to 0.215 (1%) on the left side and from 0.217 to 0.195 (10%) on the right side for $s\bar{s}$ breaks (top); $K^{\pm}$ (middle) and η, η' (bottom), using the binomial reweighting formula.

4.2 Multinomial ρ Reweighting

Figure 12 shows reweighted histograms for the ρ parameter using the updated weight, accounting for the rejection probability from eq. (19). Again ρ is decreased by 1% and 10%. Comparing these results to the previous ones show that:

1% reweighting shows similar results as before, with p-value at around 10^{-1} for the $s\bar{s}$ plots and lower for the meson plots. The 10 % reweighting is vastly improved for the string break reweighting which is reflected in the ratio plot. The meson reweighting is still not exact however, which suggests an unknown source of error in the assumptions of the transition from string breaks to mesons.

4.3 Multinomial $p_{\eta,A}$ Reweighting

Next, the η accept parameter p_{η_A} is varied. The reweighted histograms are shown in figure 13. The reweighting is increased to 10% and 20% of the original value to account for the lower impact of the parameter on the final distribution. This will however lower the overall accuracy as the parameter is further from its original value. A transition from η, η' to all diagonal mesons is made as the relative probabilities between the diagonal mesons vary when altering η_A and η'_A. The reweighting formula assumes the individual meson probabilities stay the same as discussed in section (3.3). Considering all diagonal mesons solves this problem.

What can be immediately noticed is that due to the lower difference between test and target plot, the ratio scale has been adjusted. This again lowered difference will increase p-values inherently. $s\bar{s}$ reweighting is still marginally better than the mesons. However, for kaons, this reweighting is much more accurate compared to ρ with p-values $> 10^{-20}$, while the vector mesons are very off (p-value ≈ 0). This might be due to the larger number of counted mesons, increasing the measured errors. Comparing the p-values there is an increase (10^{-4} to 10^{-2}) for the $s\bar{s}$ when increasing from 10 % to 20 % which has not been noticed before but might be statistical in nature. This suggests that reweighting $p_{\eta,A}$ is less impacted when increasing the difference between the test and target value.

4.4 Multinomial $p_{\eta',A}$ Reweighting

η'_A is altered for reweighted in figure 14. The reweighted difference is further increased to 20 % and 40 % to make differences more apparent. This again will reduce overall accuracy.

There is even smaller differences between test and target plot, and the ratio plot has been

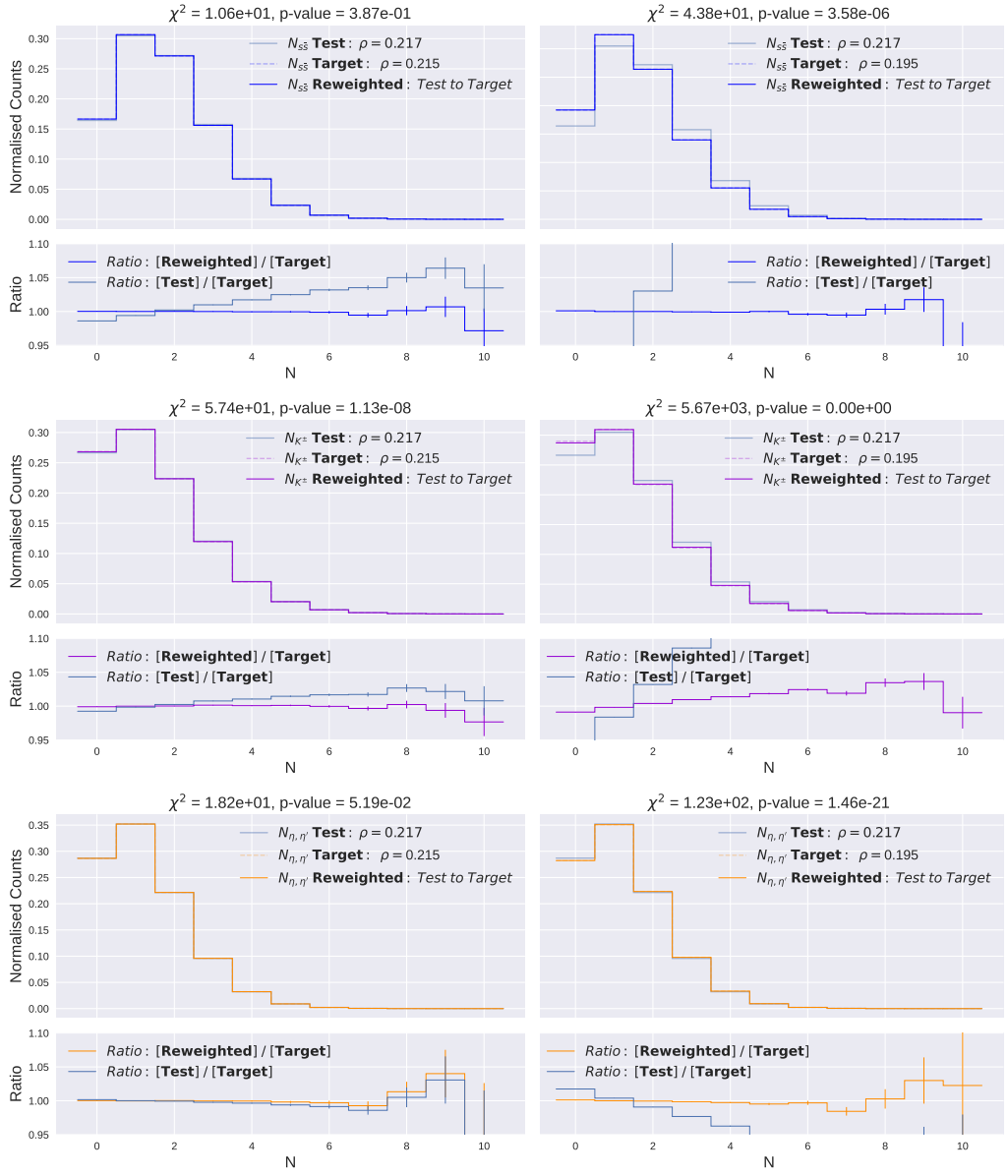

Figure 12: ρ is reweighted from 0.217 to 0.215 (1%) on the left side and from 0.217 to 0.195 (10%) on the right side for $s\bar{s}$ breaks (top); $K^{\pm}$ (middle) and η, η' (bottom), using multinomial reweighting formula.

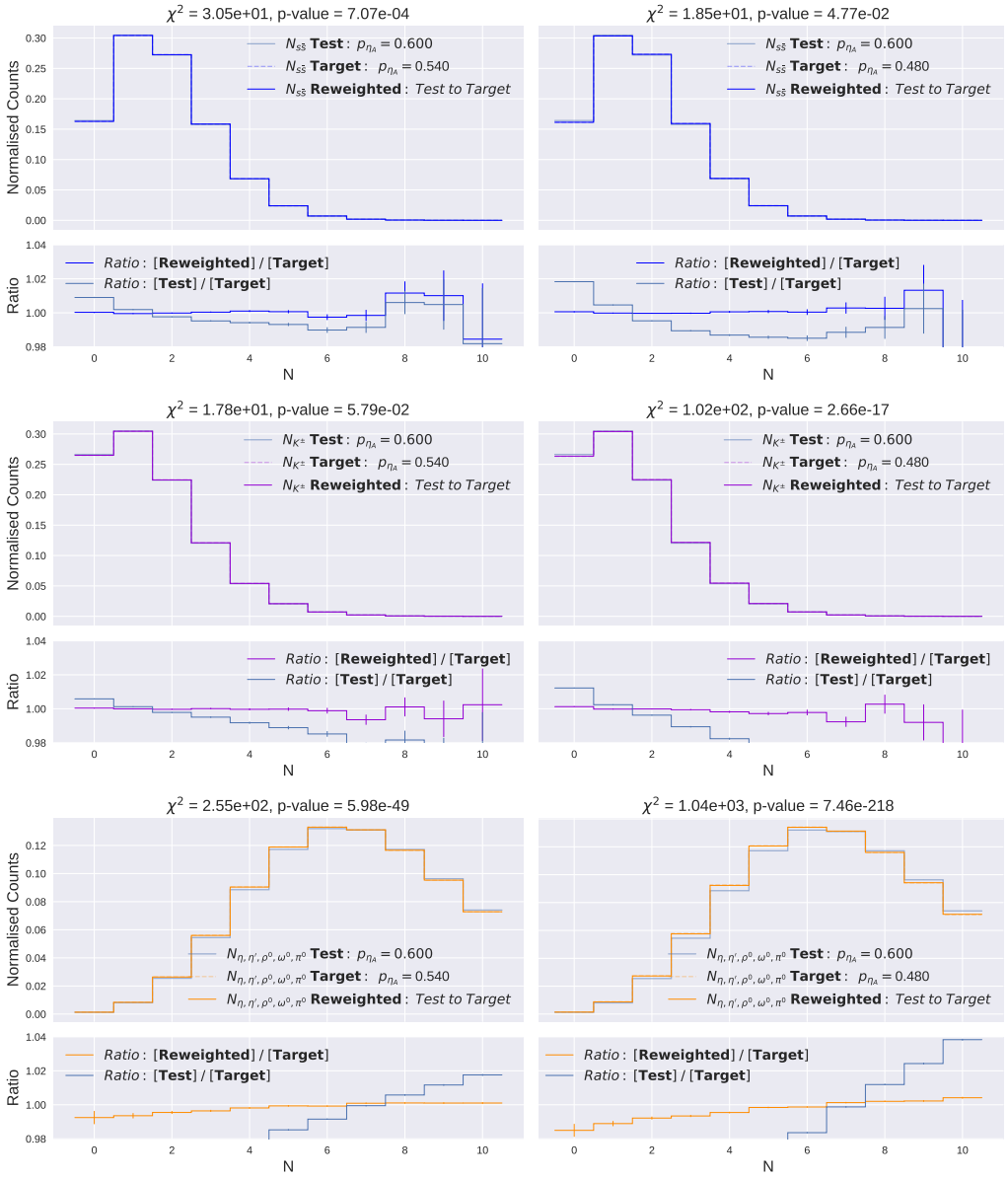

Figure 13: η_A is reweighted from 0.600 to 0.540 (10%) on the left from 0.600 to 0.480 (20%) on the right for $s\bar{s}$ breaks (top), K^+, K^- (middle) and diagonal mesons (bottom).

adjusted again. It can now be seen that statistical error is more prevalent and p-values are even more skewed. Both of these things are highlighted in the low p-value of the $s\bar{s}$ plot ($\approx 10^{-30}$) compared to $K^{\pm}(\approx 10^{-10})$ and the diagonal mesons $(\approx 10^{-20})$ for the 10 % plots. The string break reweighting should always be at least as accurate as the others. In the ratio plots, the trend from before continues: $K^{\pm}$ is very close to one for the high count N-values with low uncertainty while the vector mesons seem to have a systematic displacement.

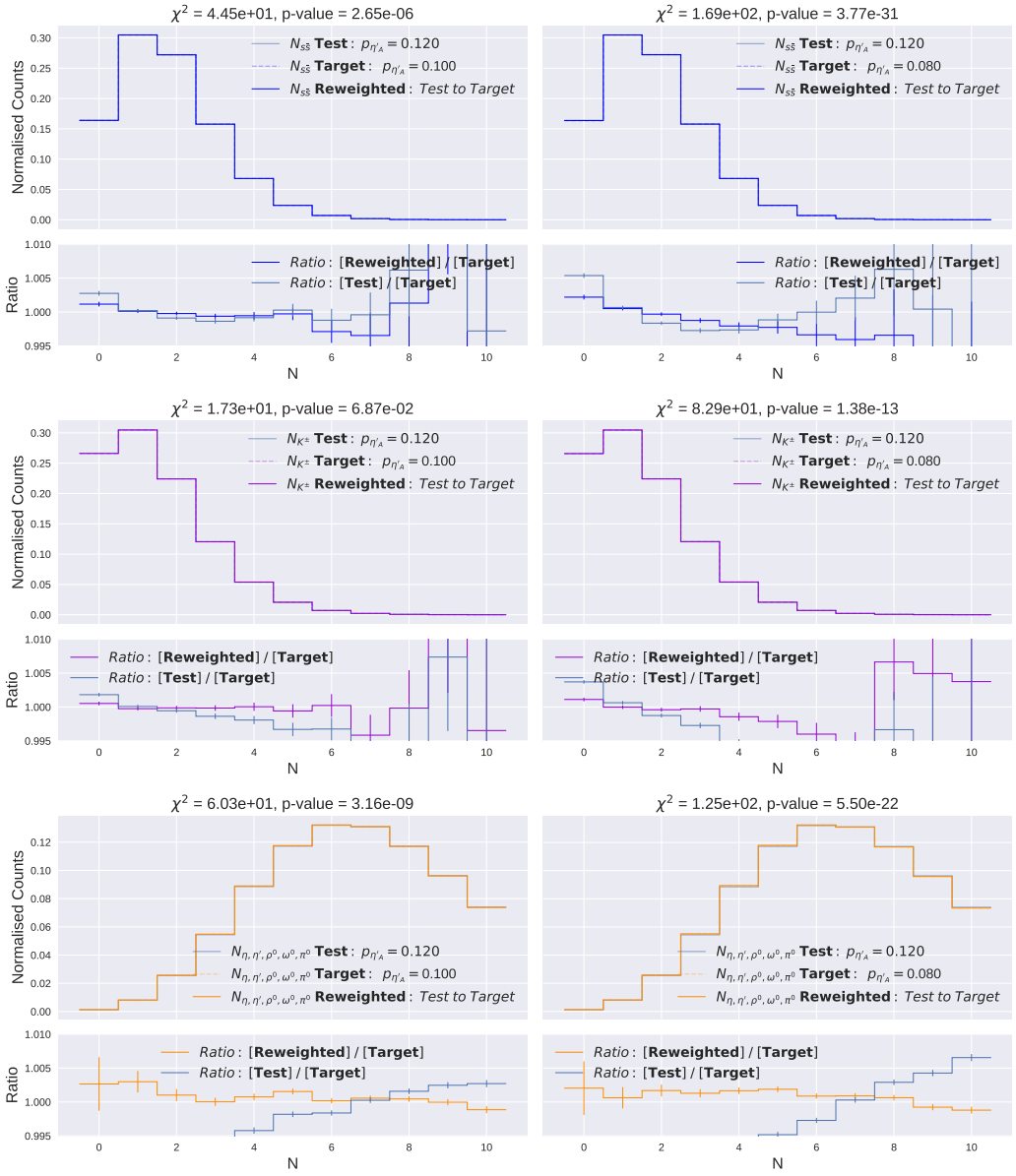

Figure 14: η'_A is reweighted from 0.120 to 0.100 (20%) on the left, from 0.120 to 0.080 (40%) on the right for $s\bar{s}$ breaks (top), K^+, K^- (middle) and diagonal mesons (bottom).

5 Conclusion

Reweighting methods allow us to take shortcuts in simulating collision events when iterating over multiple parameter values. In this thesis, reweighting methods have been implemented for three parameters: ρ, $p_{\eta,A}$ and $p_{\eta',A}$. The theory outlines the structure of the event generation and the physics behind the parameter values in the context of hadronization and Lund strings. The method continues by developing a binomial reweighting method before generalising to a more rigorous multinomial reweighting method incorporating the η and η' meson suppression. The method is used to analytically calculate weights based on transition probabilities from state diagrams. The method is implemented in PYTHIA for $e^+e^- \to Z^0 \to q\bar{q}$ collisions. The results are compared between string level and hadron level by comparing the accuracy of the reweighting $s\bar{s}$ string breaks to the reweighting of the final mesons.

The results show the relative precision of the reweighting method when considering only the string breaks. The precision dissipates when reweighting the mesons. This suggests the assumption of meson selection from string break is too simplistic. Deterministic factors such as energy conservation not considered may influence the results in deterministic ways.

To further build on this reweighting method, the results might be improved with finer approximations of the string break to meson transition, or by directly counting final mesons and reweighting based on meson probabilities. The latter one has the benefit of reweighting the property we are measuring while the former is more general and does not need to count and calculate the probabilities of each meson. Then, baryons and decays may be incorporated before comparing the method to real, experimental data. This can be done to find a finer estimate for the parameter values (tuning). One can also reweight down to the original value by biasing the final state and increasing the production and sample size of a certain meson.

This study investigates one part of reweighting in the context of event generators. As more reweighting methods are developed for other types of collisions, the toolkit is expanded further and methods might make their way into other fields. And the more simulation-based software is developed, the more need for reweighting methods. Reweighting is a natural next step in working towards more efficient simulations, where both accuracy and sustainability can increase.

6 References

[1] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: *SciPost Phys. Codebases* (2022), p. 8. DOI: 10.21468/SciPostPhysCodeb.8. URL: https://scipost.org/10.21468/SciPostPhysCodeb.8.

[2] B Andersson. *The Lund Model*. Grantor: SCOAP3. Cambridge: Cambridge University Press, 1998. ISBN: 9781009401296. DOI: 10.1017/9781009401296.

[3] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: https://www.sciencedirect.com/science/article/pii/0370157383900807.

[4] B Andersson, G Gustafson, and T Sjöstrand. "A three-dimensional model for quark and gluon jets". In: *Zeitschrift für Physik C Particles and Fields* 6 (1980), pp. 235–240. DOI: 10.1007/BF01557774. URL: https://doi.org/10.1007/BF01557774.

[5] Robert Mann. *An Introduction to Particle Physics and the Standard Model*. Boca Raton, FL: CRC Press, 2010. Chap. 16. ISBN: 978-1439810743.

[6] W.-M. Yao et al. "Review of Particle Physics". In: *Journal of Physics G* 33 (2006), pp. 1+. URL: http://pdg.lbl.gov.

[7] Theo Assiotis. *Stochastic modelling course notes*. private communication. 2023.

[8] Glen Cowan. *Statistical Data Analysis*. Oxford University Press, 1998.

Appendices

A WKB integral

To produce a quark-antiquark pair, both need to tunnel through a potential equal to the total energy of the produced particle. In 1+1 D, this is

$$\sqrt{p^2+m^2} = V = \kappa x \iff p = \sqrt{(\kappa x)^2 - m^2}, \tag{25}$$

with $0 < x < m/\kappa$.

The integral of the WKB approximation is given below for two particles

$$2\int_0^{m/\kappa} |p(x)dx| = \tag{26}$$

$$2\int_0^{m/\kappa} |\sqrt{(\kappa x)^2 - m^2}|dx = 2\int_0^{m/\kappa} |mi\sqrt{1-(\kappa x/m)^2}|dx \tag{27}$$

The square root is now real such that this becomes

$$2\int_0^{m/\kappa} m\sqrt{1-(\kappa x/m)^2}dx$$

Letting $\frac{\kappa x}{m} \to y$ we get

$$2\frac{m^2}{\kappa}\int_0^1 \sqrt{1-y^2}dy. \tag{28}$$

Again, using substitution we let $y \to \sin(\theta)$

$$2\frac{m^2}{\kappa}\int_0^{\pi/2} \sqrt{1-\sin^2(\theta)}\cos(\theta)d\theta \tag{29}$$

$$= 2\frac{m^2}{\kappa}\int_0^{\pi/2} \cos^2(\theta)d\theta$$

$$= \frac{m^2}{\kappa}\int_0^{\pi/2} 1+\cos(2\theta)d\theta$$

$$= \frac{m^2}{\kappa}\left(\frac{\pi}{2}\right) = \frac{\pi m^2}{2\kappa}$$

and thus a transmission probability

$$T \propto e^{-2\int |p(x)|dx} = \exp\left(-\frac{\pi m^2}{\kappa}\right) \tag{30}$$

B String break probabilities

The string break probabilities in eqs. (20), (21) are derived below using the first passage methods introduced in eq. 16, defining A to be the desired final state and B any of the other final states.

Starting with an s quark, the simplest outcome is ending at state $u\bar{u}d\bar{d}$ with probability $p_{s,ud}$. Using h_i to be the probability of starting at state i and ending at state $u\bar{u}d\bar{d}$ we can use the state diagram of 9 to determine the probability

Starting from state s there is a probability h_s of reaching $u\bar{u}d\bar{d}$

$$h_s = \frac{2}{2+\rho} + \frac{\rho}{2+\rho} h_{s\bar{s}}$$

Following the diagram to the next step, starting from state $s\bar{s}$ there is a probability $h_{s\bar{s}}$ of reaching $u\bar{u}d\bar{d}$.

$$h_{s\bar{s}} = p_{s,J=0} h_{s_{J=0}}$$

Continuing, solving for each h_i

$$\begin{aligned} h_{s_{J=0}} &= \sin^2\alpha h_{\eta'} + \cos^2\alpha h_\eta \\ h_{\eta'} &= (1 - p_{\eta'_A}) h_s \\ h_\eta &= (1 - p_{\eta_A}) h_s \end{aligned}$$

With this system of equations, we can solve for h_s

$$\begin{aligned} \implies h_s &= \frac{2}{2+\rho} + \frac{\rho}{2+\rho} p_{s,J=0}(\sin^2\alpha(1-p'_\eta)h_s + \cos^2\alpha(1-p_\eta)h_s) \\ \implies h_s(1 - \frac{\rho}{2+\rho} p_{s,J=0}(\sin^2\alpha(1-p'_\eta) + \cos^2\alpha(1-p_\eta))) &= \frac{2}{2+\rho} \\ \implies h_s &= \frac{\frac{2}{2+\rho}}{1 - \frac{\rho}{2+\rho} p_{s,J=0}(\sin^2\alpha(1-p'_\eta) + \cos^2\alpha(1-p_\eta))} \\ &= \frac{2}{2+\rho - \rho p_{s,J=0}(\sin^2\alpha(1-p'_\eta) + \cos^2\alpha(1-p_\eta))} \\ &= \frac{2}{2+\rho(1 - p_{s,J=0}(\sin^2\alpha(1-p'_\eta) + \cos^2\alpha(1-p_\eta))}, \end{aligned} \tag{31}$$

and thus eq. 20. There are only two outcomes, so unitarity, $p_{s,s}$ is then $1 - p_{s,ud}$, thus 21.